

This is an Accepted Manuscript of an article published by Taylor & Francis in Small GTPases on 26th Sept 2021, available online: <https://www.tandfonline.com/10.1080/21541248.2021.1984829>

INTERACTOME AND EVOLUTIONARY CONSERVATION OF DICTYOSTELID SMALL GTPASES AND THEIR DIRECT REGULATORS

Gillian Forbes¹, Christina Schilde¹, Hajara Lawal¹, Koryu Kin^{1,2}, Qingyou Du¹, Zhi-hui Chen¹, Francisco Rivero^{3*} and Pauline Schaap^{1*}

¹School of Life Sciences, University of Dundee, MSI/WTB complex, Dundee DD15EH, UK.

²Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain

³Centre for Atherothrombosis and Metabolic Disease, Hull York Medical School, Faculty of Health Sciences, University of Hull, HU67RX, Hull.

*Corresponding authors: e-mail: p.schaap@dundee.ac.uk; francisco.rivero@hyms.ac.uk

Running title: Small GTPases and their GEFs and GAPs

Keywords: GTP binding proteins, guanine nucleotide exchange factors, GTPase activating proteins, stage- and cell-type specific transcriptomics, hierarchical clustering, interactome, vesicle trafficking, cell signaling, cytokinesis, cell motility

Total work count: 9065

ABSTRACT

The GTP binding proteins known as small GTPases make up one of the largest groups of regulatory proteins and control almost all functions of living cells. Their activity is under respectively positive and negative regulation by guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs), which together with their upstream regulators and the downstream targets of the small GTPases form formidable signaling networks. While genomics has revealed the large size of the GTPase, GEF and GAP repertoire, only a small fraction of their interactions and functions have yet been experimentally explored. Dictyostelid social amoebas have been particularly useful in unraveling the roles of many proteins in the Rac-Rho and Ras-Rap families of GTPases in directional cell migration and regulation of the actin cytoskeleton. Almost completely assembled genomes and cell-type specific and developmental transcriptomes are available for *Dictyostelium* species that span the 0.5 billion years of evolution of the group from their unicellular ancestor. In this work, we identified all GTPases, GEFs and GAP from genomes representative of the four major taxon groups and investigated their phylogenetic relationships and evolutionary conservation and change in their functional domain architecture and in their developmental and cell-type specific expression. We performed hierarchical cluster analysis of the expression profiles of the ~2000 analysed genes to identify putative interacting sets of GTPases, GEFs and GAPs, which highlighted sets known to interact experimentally and many novel combinations. The work represents a valuable resource for research into all fields of cellular regulation.

INTRODUCTION

Small GTPases are major regulators of cellular function throughout the eukaryote domain. Also known as small GTP binding proteins or small G-proteins, they bind GTP and hydrolyse it to GDP and undergo a major conformational change when alternating between GTP and GDP bound states. This conformational switch serves to alter the activity of the effector proteins that they interact with. Small GTPases are generally activated by guanine nucleotide exchange factors or GEFs, which assist the exchange of bound GDP for GTP, and inactivated by GTPase activating proteins or GAPs, which activate the usually low intrinsic GTPase activity of the small GTPases.

The small GTPases are subdivided into four major and some minor families that each have their own GEFs and GAPs, although cross-regulation by GEFs and GAPs from other families also occurs. In alphabetical order, these families are the Arf-Sar GTPases with a range of functions in membrane trafficking, motility and gene expression (Sztul et al. 2019), the Rab-Ran family, with Rabs as major regulators of all aspects of endosomal membrane trafficking (Homma et al. 2020) and Rans as regulators of transport between the nucleus and cytosol (Boudhraa et al. 2020), the Rho-Rac family, major regulators of actin remodeling (Hall 2012) and the Ras-Rap family with major roles in cell proliferation and cell adhesion (Shah et al. 2019). Misregulation of members of each of the small GTPase families are a major cause of cancer and a range of other diseases, and the mammalian small GTPases have therefore been intensively studied over the past 30 years.

Many members of each of these families are also present in *Dictyostelium discoideum* (*Ddis*), an organism in Amoebozoa that is unicellular in its proliferative feeding stage, but aggregates to form migrating slugs and spore bearing fruiting structures when starved. With its ease of culture, well developed strategies for forward and reverse genetics and broad repertoire of protein imaging strategies and cell-biological and molecular techniques, it has become a popular model for studying most processes that are regulated by small GTPases (Eichinger and Rivero 2013). Processes common to most eukaryotes such as motility, cell division, phagocytosis and response to infection can be studied in the feeding stage. However, the multicellular phase of its life cycle with its chemotaxis-driven aggregation and morphogenesis, intercellular signaling and cell adhesion, autophagy, cell wall biosynthesis, cell differentiation and programmed cell death requires the repertoire of small GTPase-mediated processes to its fullest extent. The well-orchestrated succession of morphologies and behaviours during its developmental program eminently lends itself to identification of a broad range of mutants in small GTPase regulated processes. Such mutants are predominantly non-lethal, because they impair multicellular development without affecting unicellular proliferation. Studies using *Ddis* have made fundamental contributions to our knowledge on the roles of small GTPases in cell migration (Rivero and Xiong 2016; Pal et al. 2019), particularly highlighting the role of Ras activated PI3 kinase in local actin polymerization and actin wave formation during chemotaxis, cytokinesis, phagocytosis and macropinocytosis (Li et al. 2018; Gerisch et al. 2019; Williams et al. 2019). The family of Roco GTPases was initially identified in *Ddis* and then found to be widely distributed throughout eukaryotes (Wauters et al. 2019). However, despite these advantages only 24% of the 164 *Ddis* small GTPases and their 130 GEFs and 115 GAPs have as yet been functionally analysed.

While experimental strategies, such as pull-down of proteins or organelles with appropriately tagged “bait” proteins, followed by mass-spectrometric identification of the “catch” are excellent methods for identification of interacting proteins, they are expensive to perform on the entire range of GTPases. At a fraction of the cost of the experimental approach, bioinformatic analysis of co-regulated expression and evolutionary co-conservation of proteins can provide hints of

putative interactions between proteins and their involvement in conserved processes. Conversely, alterations in developmental expression or functional domain architectures between orthologous proteins may point to molecular changes that gave rise to phenotypic innovations.

Molecular phylogenetics divides Dictyostelia into four major taxon groups with *Ddis* residing in group 4 (Schaap et al. 2006; Schilde et al. 2019). Groups 1-3 consist of species that form relatively small clustered or branched fruiting bodies with maximally two cell types. Many species in these groups have retained encystation of individual amoebas, the ancestral amoebozoan strategy to survive starvation, in addition to sporulation in fruiting bodies. The group 4 species form large and robust fruiting bodies with up to three additional cell types. Their slugs show extensive migration, but as a group they have lost the ability to encyst (Romeralo et al. 2013; Schilde et al. 2014).

High quality genomes as well as developmental and cell-type specific transcriptomes for representatives of each of the four taxon groups of Dictyostelia are available (Eichinger et al. 2005; Heide1 et al. 2011; Sucgang et al. 2011; Gloeckner et al. 2016; Kin et al. 2018; Parikh et al. 2010). The presence of small GTPases in several of these genomes has already been investigated (Eichinger et al. 2005; Wilkins et al. 2005; Vlahou and Rivero 2006; Heide1 et al. 2011; Sucgang et al. 2011). However, these studies did not incorporate all or most genomes nor all GTPase families with their GAP and GEF regulators, and no transcriptome data.

In this study, we comprehensively investigated conservation and change in the presence, domain architecture, developmental regulation and cell type specificity of all GTPases and their GEFs, GAPs and other direct regulators across the four groups of Dictyostelia. We used this information to associate functions of individual GTPases, GAPs and GEFs with specific cell types and developmental stages, and to identify evolutionary trends in gene gain and loss and changes in developmental regulation in the different families of small GTPases and their regulators. For the vast majority of GTPases, the controlling GAPs and GEFs are unknown. We therefore performed a hierarchical cluster analysis of the transcriptome data to establish an “interactome” of similarly expressed GTPases, GEFs and GAPs to guide experimental studies in this major group of signaling proteins.

RESULTS

Identification of GTPases, GAPs and GEF across five dictyostelid genomes

The previous studies of GTPases and their regulators did not involve all group-representative genomes, while several families were not studied at all or in great depth. The group representative genomes used in this work are those of *Dictyostelium fasciculatum* (*Dfas*, group 1), *Polyspondylium pallidum* (*Ppal*, group 2), *Dictyostelium lacteum* (*Dlac*, group 3) and *Dictyostelium discoideum* (*Ddis*, group 4), which are all high quality, almost fully assembled genomes, and *Dictyostelium purpureum* (*Dpur*, group 4) a draft genome, which, like the other genomes, is accompanied by a developmental transcriptome. We generated Interpro scans (Jones et al. 2014) of all genomes and isolated GTPases, GAPs, GEFs and other regulators from all families by their Interpro identifiers. After construction of pilot phylogenetic trees, this initial survey was followed up by extensive BLASTp and tBLASTn searches of species proteomes and genomes to identify any missing genes (see Methods and Legends of the annotated trees in supplemental figures S1-S16 for details). The total number of identified genes in each family is listed in Table 1. A few additional genes were identified in families that were previously analysed, but for several families or species, data were not previously available.

Final phylogenetic trees were computed by Bayesian inference (Ronquist and Huelsenbeck 2003) from alignments of the isolated signature domain sequences of each family (Figs. S1-S16). The trees were annotated with the functional domain architecture of the proteins and

with heatmaps of the standardized developmental- and cell type specific expression levels of the genes, as outlined in the example tree for the Roco GTPases in figure 1. For instance, *roco6* and *pats1* in clades 4 and 5 of this tree are both upregulated in early development in all five species, but show no marked preferential expression in prestalk or prespore cells, and are not consistently upregulated in encystation, or expressed in stalk, spore or cup cells in the fruiting body. Across the other *roco* genes, developmental expression is not well conserved.

In addition to the annotated phylogenies of all GTPases, GAP and GEF families across the major groups of Dictyostelia, the Supplemental Materials section also contains a brief description of generalized roles of each family across all eukaryotes and a referenced summary of the established functions of specific family members in *Ddis*. The annotated phylogenies should prove useful to researchers to identify well-conserved genes for which the developmental regulation and cell-type specificity profile suggests that they are likely involved in a stage or cell-type specific function. The presence of functional domains in addition to the signature GTPase, GAP or GEF domains, such as protein kinase domain or SH3 domains, leucine rich repeats, WD40 repeats and pleckstrin-homology domains provides further hints of additional signaling activity or specific protein-protein interactions that are associated with the protein of choice.

All data on the presence of GTPases and their regulators across species and the conservation of their functional domains and their developmental and cell type specific expression profiles was compiled in supplemental spreadsheets Table S1.xlsx and Table S2.xlsx, respectively, and is presented in summary form in figures 2 and 3.

Patterns and phylogenetic distribution of gene and gene feature conservation across GTPase families

The compiled data on conservation and change in genes, functional domains and gene expression of the GTPases and their regulators across species also allows identify evolutionary trends, taxon group specific trends in changes in these features and trends in cell-type specificity stage of development at which members of specific families tend to be expressed. When compared with phenotypic differences between taxon groups, such data may eventually provide hints how molecular change in this group of regulators may have given rise to phenotypic innovation, while the preferential expression of specific families at some stages or cell types provides hints of their involvement in processes unique to that stage- or cell type on one hand and assist to more clearly define the functional role of the cell type on the other.

For each recorded feature, we calculated the distribution of the different states of that feature across the individual GTPase, GEF and GAP families and all combined GTPases, GEFs or GAPs (Figures 4 and 5). The most striking difference between the GTPases on one hand and the GEFs and GAPs on the other is that the GEFs and GAPs are generally well conserved across all five dictyostelid genomes, while the GTPases of all four large GTPase families show extensive species- or taxon group-specific gene amplification (compare figures 2 and 3). Only Roco GTPases and other small families are better conserved. Overall, only 26% of GTPases are conserved across all species with 68% being unique to one species (Fig. 4A). The gene amplification occurred across all five sequenced genomes, but most extensively in *Ddis* (Fig. 4B). For the GEFs and GAPs conservation across all five genomes are 79% and 83% respectively. The different families of GEFs and GAPs do not seem to have undergone equal gene amplification across species, but due to the low number of amplified genes, the differences may reflect stochastic variation.

Apart from the Roco family, the functional domain architecture of GTPases is over 90% conserved (Fig. 4C), which likely reflects that most GTPases only consist of their signature

GTPase domain, while the Roco GTPases have many other domains. The GEFs and GAPs are also more domain-rich and show overall 57% and 66% conservation of domains across species, respectively. The altered domain architectures mostly affect only 1 species or occur scattered across the phylogeny. The range of functional domains that are associated with GEFs or GAPs is broad. However, most common are domains that are involved in protein-protein or protein-lipid or phospholipid interactions, such as ANK, LRR, PQQ and WD40 repeats, zinc-finger domains (Znf_CH2, RING, LIM, ZnF_RBZ, zf-MYND, BRCT, TIR, FYVE, BBOX) and other domains (SH3, CH, IQ, GRAM). In addition, domains related to ubiquitin-mediated proteolysis (FBOX, DUSP, UBA_4, FBD) and protein kinase domains are also often found.

The developmental regulation of GTPases, GEFs and GAPs is only conserved across 35%, 31% and 28% of genes, respectively. There are also relatively large contributions of differentially regulated genes that affect only a single species or species scattered across the phylogeny. In the cases that do show clade-specific differences, those in which group 4 is different from groups 1, 2 and 3 are more frequently observed than those where the two more distantly related branches I and II show different developmental regulation (compare the size of the yellow and green bars in Fig. 4D).

When comparing specific developmental profiles, respectively 15%, 8% and 22% of GTPases, GEFs and GAPs are constitutively expressed, while 20%, 22% and 8% of each are only expressed during growth. The remaining 65-70% of genes are developmentally upregulated, with the GEFs and GAPs mostly being upregulated soon after starvation, and the GTPases showing about equal early upregulation or a peak of upregulation in mid-development (Fig. 5A). About half of all GTPases, GEFs and GAP are equally expressed in prestalk and prespore cells, while of the remainder twice more genes are expressed in prestalk than prespore cells. Exceptions are the combined small families (Gpn, Miro, Rag, Rol) of GTPases, which show prespore-specific expression, and the Arf GEFs, the Rab GEFs and GAPs, and the RCC1 proteins, which are preferentially expressed in prespore cells (Fig. 5B). Roughly half of the GTPases, GEFs and GAPs are not preferentially expressed in spores, stalk, cup or growing cells. Otherwise, the Rac-Rho and Ras-Rap GTPases and their GEFs and GAPs show preferential expression in the prestalk-derived stalk and cup cells, the Arf-Sar GTPases and their GEFs and GAPs are preferentially expressed in spores, while the Rab-Ran group show a small preference for expression in the prestalk derived stalk and cup cells (Fig. 5C). Rather surprisingly, the majority of GAPs and GEFs are upregulated in *Ppal* encystation, while the GTPases are mostly down-regulated or constitutively expressed (Fig. 5D).

Hierarchical clustering of GTPases, GEFs and GAPs according to their transcription profiles.

A subset of GTPases have been experimentally associated with regulation by specific GEFs or GAPs (Table S3.xlsx) but for the greater majority, the regulatory interactions between GTPases, GEFs and GAPs are unknown. Genes that act together in a complex can be expected to be expressed at the same developmental stage or in the same cell type. Shared developmental and cell type specific transcription profiles therefore provide information on putative protein interactions and identify yet unknown interactors. We applied hierarchical cluster analysis to associate GAPs and GEFs and other direct regulators with GTPases within and across all families.

The transcription data that were collected for this study were arranged in a linear array for orthologous genes across species, see Table S4.xlsx, sheet "Basic data". This sheet also collates the data of Table S3.xlsx on the experimentally established interactions between GTPases and their GEFs, GAPs and other regulators, the subcellular localization of the

proteins, the effects of genetic lesion on cellular function, protein association with organelles as determined by proteomics, and transcriptional responses to different food bacteria.

Hierarchical clustering was performed on the full stage- and cell type specific transcription profiles of the five *Dictyostelium* species, and on subsets thereof, comprising data from either Branch II (*Ddis*, *Dpur* and *Dlac*), Group 4 (*Ddis* and *Dpur*) and *Ddis* only. Distances between profiles were calculated by Pearson correlation and the average linkage algorithm was initially used to infer the trees. Using the Vlookup function in Excel (Table S4.xlsx, sheet Vlookup), the transcription profiles and the information on protein interactions, localization and function were re-organised to match the gene ordering in the hierarchical trees and copied to sheets “Fullprofile”, “BranchII”, “Group4”, “Ddisonly” and “Group4Complete”, with the latter inferred by “complete linkage”. Figure 6 shows the tree inferred from the Group 4 gene expression data, combined with the ordered expression profiles. Its large size precluded per gene annotation for an A4 size figure, but instead we show networks of established primary interactions for each cluster that contained such interactions. The same tree annotated for each gene with locus tags, primary and secondary interactors (other interactors of the primary partner), as well as the function and localization of the cognate proteins is shown in Figure S17.pdf.

At 75% branch height, the Group4 and BranchII based trees each separated in 15 clusters, while the *Ddisonly* and the Fullprofile based trees separated in 8 and 28 clusters, respectively. Here the height thresholds were altered to yield a similar number of clusters as the other trees. As a means to discriminate between the efficacy of the five clustering analyses to identify interacting proteins, we counted the total number of experimentally established primary and secondary interactions in each cluster. Most (85) interactions were recovered in the Group4 based tree, but the others followed closely behind with 83 or 82 interactions (Table S5). The use of complete linkage clustered only 59 interactions. The trees that were inferred from the different subsets of the data showed different topologies, but comparison of their cluster content revealed that genes that clustered together in one tree often also clustered together in the other (Figure S18.pdf). This was particularly the case when comparing the Group4 with the BranchII and *Ddisonly* trees, but less so with the Fullprofile based tree. However, even the latter tree had over 50% of its nodes in common with the other trees (Figure S19).

The purpose of the transcriptome based cluster analysis is to identify novel interacting networks of GTPases, GEFs and GAPs. Figure S18.pdf can assist in identifying groups of genes that robustly cluster together, even when different subsets of the transcriptome data are used. Further identification of novel gene interactions can also be guided by the statistical support for the tree nodes that combine genes together. The trees were therefore subjected to approximately unbiased (AU) and selective inference (SI) bootstrapping (Suzuki and Shimodaira 2006; Shimodaira and Terada 2019) (Figure S20.pdf). The deeper nodes that separate the larger clusters show poor statistical support, but many smaller clusters closer to the tips of the branches are moderately to well supported.

DISCUSSION

We investigated conservation and change in the small GTPases and their GEFs, GAPs and other direct regulators across the four major taxon groups of Dictyostelia and examined trends in the evolution of their developmental and cell-type specific expression and functional domain architecture. In addition, we performed a hierarchical cluster analysis of the gene expression data to identify sets of genes with similar expression profiles that potentially interact as proteins.

The phylogenetically ordered separate families of all small GTPases, GEFs and GAPs, annotated with developmental expression and functional domains (Figs. S1-S16) provides researchers with a complete inventory of all small GTPases and their regulators across

Dictyostelia. Conservation across taxon groups of proteins of interest provides a clue to their potential involvement in core regulatory processes. Furthermore, changes in the presence of genes, their developmental regulation or their functional domains yield hints of how their function may have changed in the course of evolution, which together with gene knock-out and gene replacement across species can provide information on how molecular change in this important group of cellular regulators gave rise to phenotypic innovation.

Small GTPases underwent massive gene amplification, but their regulators did not

Total numbers of small GTPases in the different families are somewhat higher in taxon group 4 but otherwise similar across Dictyostelia, while the different GEF and GAP families show similar numbers across all taxon groups. For the GEFs and GAPs, this marks their almost complete conservation as orthologs across all taxon groups, but this is not the case for the small GTPases, which underwent extensive gene amplification in individual group-representative species (Figs. 2 and 4). A eukaryote wide study of the Ras-Rap family did not particularly highlight this difference in conservation between GTPases and their regulators (van Dam et al. 2011). Extensive species-specific expansions of Rab GTPases were observed in Amoebozoa and Excavates (Brighthouse et al. 2010; Porfirio-Sousa et al. 2021), but GEFs and GAPs were not studied in parallel.

The amplification of GTPase genes in Dictyostelia may mark a species- and/or niche-specific demand for a larger number of small GTPases e.g. to be able to recognize and consume a larger variety of food bacteria or to respond to niche-specific predators, infectious agents or toxins. Some amplified genes are not or poorly expressed under laboratory conditions, which either reflects that the genes are not functional, or only expressed under specific conditions. However, the expression profiles of the amplified unique genes are not markedly different from those of the conserved small GTPases (Table 2), indicating that gene amplification did not occur in response to a single stage-specific challenge. The underlying cause for the GTPase gene amplification is therefore unclear.

Validity of the newly uncovered putative GAPs and GEFs

The GEFs, GAPs and other regulators of the *Ddis* Rac-Rho and Ras-Rap families were mostly already previously identified (Eichinger et al. 2005; Wilkins et al. 2005; Vlahou and Rivero 2006; Heidel et al. 2011; Sugang et al. 2011), and many have been functionally studied by gene manipulation and other experimental approaches (see (Plattner 2013; Rivero and Xiong 2016; Pal et al. 2019; van Egmond and van Haastert 2010). However, in the Arf-Sar family, a biological role has only been established for ArfA and for two ArfGEFs and ArfGAPs (Shina et al. 2010; Dias et al. 2013; Muller et al. 2013; Bailo et al. 2014), while other GEFs and GAPs were readily identified by their conserved Sec7 and ArfGAP domains, respectively. The Rab-Ran GTPases were least studied in *Ddis*, and their GEFs and most of their GAPs were not previously identified. Metazoan RabGAP activity resides in the well-conserved TBC domain and 31 well-conserved proteins with this domain are present across dictyostelid genomes (Fig. S6). For two of those, Dsgg and CnrF, RabGAP activity on Rab8A and Rab11A, respectively, was demonstrated (Du et al. 2008; Essid et al. 2012; Parkinson et al. 2014).

In Metazoa, RabGEF activity is associated with a wide range of structurally divergent proteins, such as the ~12 subunit TRAPP (TRANsport Protein Particle) I, II, III complexes, which are conserved throughout eukaryotes (Kim et al. 2016) and proteins with VPS9 and DENN domain. We identified 8 and 18 well conserved DENN and VPS9 domain proteins in Dictyostelia, respectively, as well as 4 close homologs to other metazoan RabGEFs (Fig. S5). However, no RabGEF activity has been established for any of the *Ddis* proteins. Their

assignment as putative RabGEFs should therefore be treated with caution, but simultaneously as a wide-open field for gene function discovery.

Rac-Rho and Ras-Rap GTPase function is particularly associated with prestalk cells

The evolutionary trends in conservation of GTPases, GEFs and GAPs were unremarkable, with most genes, functional domains and expression profiles being conserved. As is the case for transcription factors, the only other large family of cellular regulators that was analysed in the same manner (Forbes et al. 2019), there is a tendency for expression profiles of GTPase related genes to be more different between taxon group 4 and the other three groups, than between the phylogenetically more distant branches I and II that contain groups 1 & 2 and groups 3 & 4, respectively (Fig. 4D). Such a trend is not found for differences in functional domains but is correlated with group 4 being phenotypically the most distinctive. As previously suggested, this could mean that changes in gene expression played a larger role in phenotypic innovation than those in functional domains.

A marked difference between transcription factors and GTPase related proteins, is that two times more transcription factor genes were expressed in prespore over prestalk cells, whereas almost the opposite is true for the GTPase related genes. For the mature cell types, both the transcription factors and the GTPase related genes show as groups slightly higher expression in stalk over spore cells. The prestalk preference is particularly high for the Roco proteins and for Rac-Rho and Ras-Rap GTPases and their GEFs and GAPs, whereas the Rab-Ran GTPases are less prestalk-enriched, and their GEFs and GAPs distinctly prespore enriched. Their prestalk preferences likely reflect the dominant role of the Rac-Rho and Ras-Rap GTPases in regulating the actin cytoskeleton in coordinated cell migration, which in slugs largely depends on the prestalk population (Chien et al. 2000; Rieu et al. 2009; Flegel et al. 2011). Both the relatively low dependence of prestalk cells on novel transcription factors and their high dependence on Rac-Rho and Ras-Rap GTPases suggest that this population is much more involved in morphogenetic cell movement than preparing itself for stalk cell differentiation.

The prespore cells are at this stage involved in the expression of spore coat genes and in laying down the first layer of the spore wall and synthesizing spore wall precursors in Golgi derived vesicles (Srinivasan et al. 1999; West 2003). Since the Rab GTPases are major coordinators of vesicle trafficking, amongst which the transport of vesicles from the endoplasmic reticulum through the Golgi apparatus (Lamber et al. 2019; Homma et al. 2020), the relative prespore enrichment of this family and its regulators may reflect the involvement of Rabs in prespore vesicle formation.

Towards elucidation of the GTPase interactome.

Proteins that act in complexes, like the GTPases and their GAPs and GEFs, have to be expressed at the same developmental stage and in the same cell type to be able to do so. Information on shared transcriptional regulation of their cognate genes can therefore help to recognize interactions between specific proteins. To this end, we subjected the extensive transcription data that we collected for all GTPase related genes to statistical procedures to identify clusters of similarly related genes. We estimated the efficacy of different clustering methods and subsets of the transcription data by examining to what extent experimentally recognized interacting proteins also clustered in our analysis. Use of only the group 4 (*Ddis* and *Dpur*) transcription data recovered most of the known interactions in clusters but use of larger and smaller subsets performed almost equally well and about half of the genes consistently clustered together. If known interacting proteins did not occupy the same cluster, their

transcription heatmaps were quite obviously different, which likely means that some interaction partners are used over a wider range of stages and cell types than their counterpart.

The known interactions are mainly between members of the Ras-Rap and Rho-Rac families and they were mostly recovered from clusters C2/C8 and C11, respectively (Figure 6). Cluster C2 contains genes that peak during aggregation and show specificity for stalk and/or cup cells. Cluster 8 contains genes with peak expression in mid development or growth and early development and show highest expression in both spores and growing cells. Cluster 11 contains genes which are downregulated after growth and early development and are mostly weakly expressed in spore, stalk and cup cells. Only a few known interactions were recovered from clusters that showed upregulation in later development. This bias likely reflects the focus of workers in the *Dictyostelium* field on directed cell migration and regulation of the actin cytoskeleton, which play a major role in aggregation, and in cytokinesis and phagocytosis in the growth stage. GTPase related proteins in these processes may yet be uncovered by scrutiny of other members of clusters C2, C8 and C11, while the developmental profiles of the remaining clusters may hint to requirements of specific GTPases in slug and fruiting body morphogenesis, and processes like phototaxis and thermotaxis.

In addition, there is still a wide-open field of discovery for elucidation of functions for the Rab-Ran and Arf-Sar families in processes like phagocytosis, autophagy and cell wall biosynthesis that have a strong stage- and/or cell type specific component. With respect to the Ran GTPases, cluster 14 is of interest. This cluster harbors just 6 genes but 5 of them are related to proteins of the Ran pathway, including the two Ran homologs, RanA and RanB, RanBP1 and two uncharacterized genes, DDB_G0278125 and DDB_G0269700, encoding proteins with RCC1 domains characteristic of Ran GEFs. RanBP1 is well conserved across eukaryotes and acts as a guanine nucleotide dissociation inhibitor and as a RanGAP cofactor. Interestingly, both RanA and RanBP1 localize to the bacteria-containing vacuole upon infection with *Legionella* (Hilbi et al. 2014), further strengthening a possible physical interaction. Of the RCC1 proteins, the first has no obvious orthologs in higher eukaryotes but the second shares domain architecture with human RCBTB1, a protein mutated in a rare form of retinopathy for which an effect on Ran was postulated (Solomou et al. 2003; Coppieters et al. 2016). These two RCC1 proteins emerge as strong candidates for future studies aiming at identifying Ran GEFs in *Ddis*.

On the other hand, the biological roles of many of the 400 proteins described here may be relatively minor or overlapping with other family members. Furthermore, important constitutively expressed GEFs or GAPs may interact with several GTPases that show more stage- and cell type specific expression profiles or *vice versa*, in which cases their interactions are not evident from the cluster analysis. The results of this analysis should therefore be interpreted and used with caution and preferably only to substantiate information acquired from other approaches.

METHODS

GTPase sequence retrieval

Small GTPases were isolated from Interpro scans (Jones et al. 2014) of the most recently annotated proteomes of *Ddis*, *D. purpureum* (*Dpur*) (<http://dictybase.org/Downloads/>), *D. lacteum* (*Dlac*), *Polysphondylium pallidum* (*Ppal*) and *D. fasciculatum* (*Dfas*) (<http://sacgb.leibniz-fli.de/cgi/index.pl?ssi=free>) using the InterPro identifiers IPR001806, IPR002041, IPR003578, IPR020849, IPR021181 and IPR006689 for generic, Ran, Rho, Ras, Miro and Arf GTPases respectively. Because of considerable sequence similarity overlap between different categories of GTPases, a single sequence alignment was made of all

retrieved proteins using Clustal Omega (Sievers and Higgins 2014) with five combined iterations, and a pilot tree was inferred by RaxML (Stamatakis 2014). This tree subdivided the sequences into 4 large groupings, which each mostly contained members of Arf-Sar, Rab-Ran, Rac-Rho or Ras-Rap families of GTPases, respectively, and some smaller clades that included the Roco, Miro, Rag, Gpn and Roco-like GTPases. BLASTp searches were performed with clade-representative members within each of the groupings to retrieve any genes that were not detected by the InterPro scans. New alignments were prepared for each of the major and minor groups, and phylogenies were inferred by Bayesian analysis (Ronquist and Huelsenbeck 2003). These trees were scrutinized for any missing members of otherwise orthologous sets, and further BLASTp or tBLASTn searches were performed with a member of the set as bait. Final Bayesian trees were then inferred including the additional hits, using a mixed amino-acid model with rate variation between sites predicted by a gamma distribution. Analyses were run for 1 to 10 million generations but did often not fully converge due to the relatively small number of variable sites in the GTPase domains. Orthologous clades were generally well resolved but deeper nodes of the trees were not.

Sequence retrieval for GTP-ase regulators

The functional domains of GEFs, GAPs and other regulators of small GTPases are distinctive, and thus far specific for the different subtypes of GTPases. These proteins were identified in a first round from InterPro scans using the InterPro identifiers of their GTPase regulatory domain as indicated in the legends to figures S2-S15. After inferring a pilot tree for each regulator type, BLASTp and tBLASTn searches of proteomes and genomes were performed to identify any missing members of the family as described above and a final tree was inferred by Bayesian analysis.

Phylogenetic tree annotation

The functional domain architectures, including PFAM domains, signal peptides and internal repeats were analysed in SMART (Schultz et al. 1998), saved as .svg files and juxtaposed to the protein locus tags at the tips of the phylogenetic tree branches. SMART or PFAM domain identifiers are listed in each figure and domain descriptions can be retrieved from http://smart.embl-heidelberg.de/smart/domain_table.cgi or <http://pfam.xfam.org/browse>. Clades of orthologous genes or other groupings were annotated with relative transcript levels at specific developmental stages or in specific cell types, shown as heat maps that represent the fraction of the maximum transcript read count for the developmental profiles and the fraction of the summed read counts for the cell types. The normalized reads were retrieved from published RNA sequencing experiments (Gloeckner et al. 2016; Kin et al. 2018; Forbes et al. 2019; Parikh et al. 2010).

Hierarchical clustering

The full set or subsets of transcriptome data for *Ddis* genes and their orthologs in *Dpur*, *Dlac*, *Ppal* and *Dfas*, were re-ordered into a linear array and subjected to hierarchical clustering in Orange 3.27.1 (Demsar et al. 2013), generally using Pearson correlation as the distance metric and average linkage. The data from individual experiments were included for *Ddis* and *Dlac* cell-type specific transcripts, rather than the averaged values used in Figs.1 and Figs. S1-S16. Data were standardized as percentage of the maximum value of a set in general, or to the sum of values, when there were only two. Hierarchical trees were bootstrapped with 1000 bootstrap replicates using the pvclust package in R (Suzuki and Shimodaira 2006).

To compare trees obtained from different subsets of the transcription profiles, two trees were juxtaposed using the tanglegram function in the dendextend package in R (Galili 2015).

Correlations between common nodes in tree comparisons were calculated and plotted using the dendextend and corrplot packages in R.

Retrieval of data on protein interaction, function and subcellular localization

Data on protein interactions, subcellular localization and phenotypes or components of Rho signaling were collected from (Rivero and Xiong 2016). This dataset and data for other and/or more recently studied GTPases and regulators were retrieved from publications listed at the corresponding gene page in Dictybase (<http://dictybase.org/>) (Fey et al. 2013). The information is compiled in Table S3.xlsx. Networks of interacting proteins were visualized using Cytoscape (Shannon et al. 2003).

FUNDING DETAILS

This research was funded by the European Research Council under grant 742288 and by The Wellcome Trust under grant 100293/Z/12/Z.

DISCLOSURE STATEMENTS

No financial benefits have arisen from the application of this work.

DATA ACCESS

All original sequence data used in this work were previously archived in the European Nucleotide Archive and/or Genbank with earlier publications. All processed data are provided in the Supplemental spreadsheets Supplemental_Table_S1.xlsx, Supplemental_Table_S2.xlsx, Supplemental_Table_S3.xlsx and Supplemental_Table_S4.xlsx. Supplemental figures S1-S16, S19 and S20 are provided in Supplemental Materials.pdf. Supplemental figures S17, S18 and S19 are provided as separate .pdf files due to their large size.

DISCLOSURE STATEMENTS

No financial benefits have arisen from the application of this work.

AUTHOR CONTRIBUTIONS

G.F. identified and analysed all GTPases and performed cluster analysis, C.S., H.L., Q-Y.D. and Z.-H.C. identified and analyses GAPs and GEFs across species, K.K. computed phylogenetic trees, F.R. collected data on interaction, localization and function of proteins, advised on and curated all analyses and the manuscript draft. P.S. designed the study and wrote the manuscript.

REFERENCES

- Bailo N, Cosson P, Charette SJ, Paquet VE, Doublet P, Letourneur F. 2014. Defective lysosome maturation and Legionella pneumophila replication in Dictyostelium cells mutant for the Arf GAP ACAP-A. *J Cell Sci* **127**: 4702-4713.
- Boudhraa Z, Carmona E, Provencher D, Mes-Masson A-M. 2020. Ran GTPase: A Key Player in Tumor Progression and Metastasis. *Frontiers in Cell and Developmental Biology* **8**.
- Brighouse A, Dacks JB, Field MC. 2010. Rab protein evolution and the history of the eukaryotic endomembrane system. *Cellular and molecular life sciences : CMLS* **67**: 3449-3465.
- Chien S, Chung CY, Sukumaran S, Osborne N, Lee S, Ellsworth C, McNally JG, Firtel RA. 2000. The Dictyostelium LIM domain-containing protein LIM2 is essential for proper chemotaxis and morphogenesis. *Mol Biol Cell* **11**: 1275-1291.

- Coppieters F, Ascari G, Dannhausen K, Nikopoulos K, Peelman F, Karlstetter M, Xu M, Brachet C, Meunier I, Tsilimbaris MK et al. 2016. Isolated and Syndromic Retinal Dystrophy Caused by Biallelic Mutations in RCBTB1, a Gene Implicated in Ubiquitination. *Am J Hum Genet* **99**: 470-480.
- Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A et al. 2013. Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research* **14**: 2349-2353.
- Dias M, Blanc C, Thazar-Poulot N, Ben Larbi S, Cosson P, Letourneur F. 2013. Dictyostelium ACAP-A is an ArfGAP involved in cytokinesis, cell migration and actin cytoskeleton dynamics. *J Cell Sci* **126**: 756-766.
- Du F, Edwards K, Shen Z, Sun B, De Lozanne A, Briggs S, Firtel RA. 2008. Regulation of contractile vacuole formation and activity in Dictyostelium. *EMBO J* **27**: 2064-2076.
- Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q et al. 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**: 43-57.
- Eichinger L, Rivero F, ed. 2013. *Dictyostelium discoideum protocols*. Humana Press.
- Essid M, Gopaldass N, Yoshida K, Merrifield C, Soldati T. 2012. Rab8a regulates the exocyst-mediated kiss-and-run discharge of the Dictyostelium contractile vacuole. *Mol Biol Cell* **23**: 1267-1282.
- Fey P, Dodson RJ, Basu S, Chisholm RL. 2013. One stop shop for everything Dictyostelium: dictyBase and the Dicty Stock Center in 2012. *Methods Mol Biol* **983**: 59-92.
- Flegel KA, Pineda JM, Smith TS, Laszczyk AM, Price JM, Karasiewicz KM, Damer CK. 2011. Copine A is expressed in prestalk cells and regulates slug phototaxis and thermotaxis in developing Dictyostelium. *Dev Growth Differ* **53**: 948-959.
- Forbes G, Chen ZH, Kin K, Lawal HM, Schilde C, Yamada Y, Schaap P. 2019. Phylogeny-wide conservation and change in developmental expression, cell-type specificity and functional domains of the transcriptional regulators of social amoebas. *BMC Genomics* **20**: 890.
- Galili T. 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**: 3718-3720.
- Gerisch G, Prassler J, Butterfield N, Ecke M. 2019. Actin Waves and Dynamic Patterning of the Plasma Membrane. *The Yale journal of biology and medicine* **92**: 397-411.
- Gloeckner G, Lawal HM, Felder M, Singh R, Singer G, Weijer CJ, Schaap P. 2016. The multicellularity genes of dictyostelid social amoebas. *Nature communications* **7**: 12085.
- Hall A. 2012. Rho family GTPases. *Biochemical Society transactions* **40**: 1378-1382.
- Heidel A, Lawal H, Felder M, Schilde C, Helps N, Tunggal B, Rivero F, John U, Schleicher M, Eichinger L et al. 2011. Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. *Genome Res* **21**: 1882-1891.
- Hilbi H, Rothmeier E, Hoffmann C, Harrison CF. 2014. Beyond Rab GTPases Legionella activates the small GTPase Ran to promote microtubule polymerization, pathogen vacuole motility, and infection. *Small GTPases* **5**: e972859.
- Homma Y, Hiragi S, Fukuda M. 2020. Rab family of small GTPases: an updated view on their regulation and functions. *The FEBS journal* doi:10.1111/febs.15453.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* doi:10.1093/bioinformatics/btu031.
- Kim JJ, Lipatova Z, Segev N. 2016. TRAPP Complexes in Secretion and Autophagy. *Frontiers in Cell and Developmental Biology* **4**.

- Kin K, Forbes G, Cassidy A, Schaap P. 2018. Cell-type specific RNA-Seq reveals novel roles and regulatory programs for terminally differentiated Dictyostelium cells. *BMC Genomics* **19**: 764.
- Lamber EP, Siedenburg AC, Barr FA. 2019. Rab regulation by GEFs and GAPs during membrane traffic. *Curr Opin Cell Biol* **59**: 34-39.
- Li X, Edwards M, Swaney KF, Singh N, Bhattacharya S, Borleis J, Long Y, Iglesias PA, Chen J, Devreotes PN. 2018. Mutually inhibitory Ras-PI(3,4)P2 feedback loops mediate cell migration. *Proc Natl Acad Sci U S A* **115**: E9125-e9134.
- Muller R, Herr C, Sukumaran SK, Omosigho NN, Plomann M, Riyahi TY, Stumpf M, Swaminathan K, Tsangarides M, Yiannakou K et al. 2013. The cytohesin paralog Sec7 of Dictyostelium discoideum is required for phagocytosis and cell motility. *Cell Commun Signal* **11**: 54.
- Pal DS, Li X, Banerjee T, Miao Y, Devreotes PN. 2019. The excitable signal transduction networks: movers and shapers of eukaryotic cell migration. *Int J Dev Biol* **63**: 407-416.
- Parikh A, Miranda ER, Katoh-Kurasawa M, Fuller D, Rot G, Zagar L, Curk T, Sugang R, Chen R, Zupan B et al. 2010. Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biol* **11**: R35.
- Parkinson K, Baines AE, Keller T, Gruenheit N, Bragg L, North RA, Thompson CR. 2014. Calcium-dependent regulation of Rab activation and vesicle fusion by an intracellular P2X ion channel. *Nat Cell Biol* **16**: 87-98.
- Plattner H. 2013. Contractile vacuole complex--its expanding protein inventory. *Int Rev Cell Mol Biol* **306**: 371-416.
- Porfírio-Sousa AL, Tice AK, Brown MW, D JGL. 2021. Phylogenetic reconstruction and evolution of the Rab GTPase gene family in Amoebozoa. *Small GTPases* doi:10.1080/21541248.2021.1903794: 1-14.
- Rieu JP, Saito T, Delanoe-Ayari H, Sawada Y, Kay RR. 2009. Migration of Dictyostelium slugs: Anterior-like cells may provide the motive force for the prespore zone. *Cell Motil Cytoskeleton* doi:10.1002/cm.20411 [doi].
- Rivero F, Xiong H. 2016. Rho Signaling in Dictyostelium discoideum. *Int Rev Cell Mol Biol* **322**: 61-181.
- Romeralo M, Skiba A, Gonzalez-Voyer A, Schilde C, Lawal H, Kedziora S, Cavender JC, Glockner G, Urushihara H, Schaap P. 2013. Analysis of phenotypic evolution in Dictyostelia highlights developmental plasticity as a likely consequence of colonial multicellularity. *Proc Biol Sci* **280**: 20130976.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.
- Schaap P, Winckler T, Nelson M, Alvarez-Curto E, Elgie B, Hagiwara H, Cavender J, Milano-Curto A, Rozen DE, Dingeramn T et al. 2006. Molecular phylogeny and evolution of morphology in the social amoebas. *Science* **314**: 661-663.
- Schilde C, Lawal HM, Kin K, Shibano-Hayakawa I, Inouye K, Schaap P. 2019. A well supported multi gene phylogeny of 52 dictyostelia. *Mol Phylogenet Evol* **134**: 66-73.
- Schilde C, Skiba A, Schaap P. 2014. Evolutionary reconstruction of pattern formation in 98 Dictyostelium species reveals that cell-type specialization by lateral inhibition is a derived trait. *EvoDevo* **5**: 34.
- Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* **95**: 5857-5864.
- Shah S, Brock EJ, Ji K, Mattingly RR. 2019. Ras and Rap1: A tale of two GTPases. *Seminars in cancer biology* **54**: 29-39.

- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498-2504.
- Shimodaira H, Terada Y. 2019. Selective Inference for Testing Trees and Edges in Phylogenetics. *Frontiers in Ecology and Evolution* **7**.
- Shina MC, Muller R, Blau-Wasser R, Glockner G, Schleicher M, Eichinger L, Noegel AA, Kolanus W. 2010. A cytohesin homolog in Dictyostelium amoebae. *PLoS One* **5**: e9378.
- Sievers F, Higgins DG. 2014. Clustal omega, accurate alignment of very large numbers of sequences. *Methods in molecular biology* **1079**: 105-116.
- Solomou EE, Sfrikakis PP, Kotsi P, Papaioannou M, Karali V, Vervessou E, Hoffbrand AV, Panayiotidis P. 2003. 13q deletion in chronic lymphocytic leukemia: characterization of E4.5, a novel chromosome condensation regulator-like guanine nucleotide exchange factor. *Leuk Lymphoma* **44**: 1579-1585.
- Srinivasan S, Alexander H, Alexander S. 1999. The prespore vesicles of Dictyostelium discoideum. Purification, characterization, and developmental regulation. *J Biol Chem* **274**: 35823-35831.
- Stamatakis A. 2014. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* doi:10.1093/bioinformatics/btu033.
- Sucgang R, Kuo A, Tian X, Salerno W, Parikh A, Feasley CL, Dalin E, Tu H, Huang E, Barry K et al. 2011. Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. *Genome Biol* **12**: R20.
- Suzuki R, Shimodaira H. 2006. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**: 1540-1542.
- Sztul E, Chen PW, Casanova JE, Cherfils J, Dacks JB, Lambright DG, Lee FS, Randazzo PA, Santy LC, Schürmann A et al. 2019. ARF GTPases and their GEFs and GAPs: concepts and challenges. *Mol Biol Cell* **30**: 1249-1271.
- van Dam TJ, Bos JL, Snel B. 2011. Evolution of the Ras-like small GTPases and their regulators. *Small GTPases* **2**: 4-16.
- van Egmond WN, van Haastert PJ. 2010. Characterization of the Roco protein family in Dictyostelium discoideum. *Eukaryot Cell* **9**: 751-761.
- Vlahou G, Rivero F. 2006. Rho GTPase signaling in Dictyostelium discoideum: Insights from the genome. *Eur J Cell Biol* **85**: 947-959.
- Wauters L, Versées W, Kortholt A. 2019. Roco Proteins: GTPases with a Baroque Structure and Mechanism. *International journal of molecular sciences* **20**.
- West CM. 2003. Comparative analysis of spore coat formation, structure, and function in Dictyostelium. *Int Rev Cytol* **222**: 237-293.
- Wilkins A, Szafranski K, Fraser DJ, Bakthavatsalam D, Muller R, Fisher PR, Glockner G, Eichinger L, Noegel AA, Insall RH. 2005. The Dictyostelium genome encodes numerous RasGEFs with multiple biological roles. *Genome Biol* **6**.
- Williams TD, Paschke PI, Kay RR. 2019. Function of small GTPases in Dictyostelium macropinocytosis. *Philos Trans R Soc Lond B Biol Sci* **374**: 20180150.

TABLES

Table 1. Numbers of GTPases, GEFs and GAPs in group representative dictyostelid genomes

	GTP-ases					GEFs					GAPs				
Family	<i>D</i> <i>di</i> <i>s</i>	<i>D</i> <i>p</i> <i>u</i> <i>r</i>	<i>D</i> <i>l</i> <i>a</i> <i>c</i>	<i>P</i> <i>p</i> <i>a</i> <i>l</i>	<i>D</i> <i>f</i> <i>a</i> <i>s</i>	<i>D</i> <i>di</i> <i>s</i>	<i>D</i> <i>p</i> <i>u</i> <i>r</i>	<i>D</i> <i>l</i> <i>a</i> <i>c</i>	<i>P</i> <i>p</i> <i>a</i> <i>l</i>	<i>D</i> <i>f</i> <i>a</i> <i>s</i>	<i>D</i> <i>di</i> <i>s</i>	<i>D</i> <i>p</i> <i>u</i> <i>r</i>	<i>D</i> <i>l</i> <i>a</i> <i>c</i>	<i>P</i> <i>p</i> <i>a</i> <i>l</i>	<i>D</i> <i>f</i> <i>a</i> <i>s</i>
Arf-Sar	24	22	24	23	22	7	6	6	7	6	12	11	12	12	13
Rab-Ran	74	60	49	49	52	12	12	13	12	12	30	30	31	31	31
Rac-Rho	22	19	15	22	15	46	46	46	44	45	47	44	48	47	47
Elmo						5	5	6	5	7					
Dock						8	8	8	8	8					
Ras	33	19	16	23	28	29	29	27	28	27	15	15	15	18	15
Rap	3	3	4	4	4						11	11	11	11	11
Roco	11	9	9	10	10										
Gpn/Rag/ Miro/Rol	7	7	7	7	7										
Total	175	139	124	138	138	107	106	106	104	105	115	111	117	119	117

Table 2. Expression profiles of conserved and unique GTPases

GTPase s	% of genes with this expression profile					
	constitu tive	decreas e after growth	peak in mid. develop	dip in mid. develop	up in early develop	up in late develop
conserved	18.7	13.2	26.4	2.2	34.1	5.5
unique	14.4	25.5	21.6	6.5	24.8	7.2

LEGENDS (and figures)

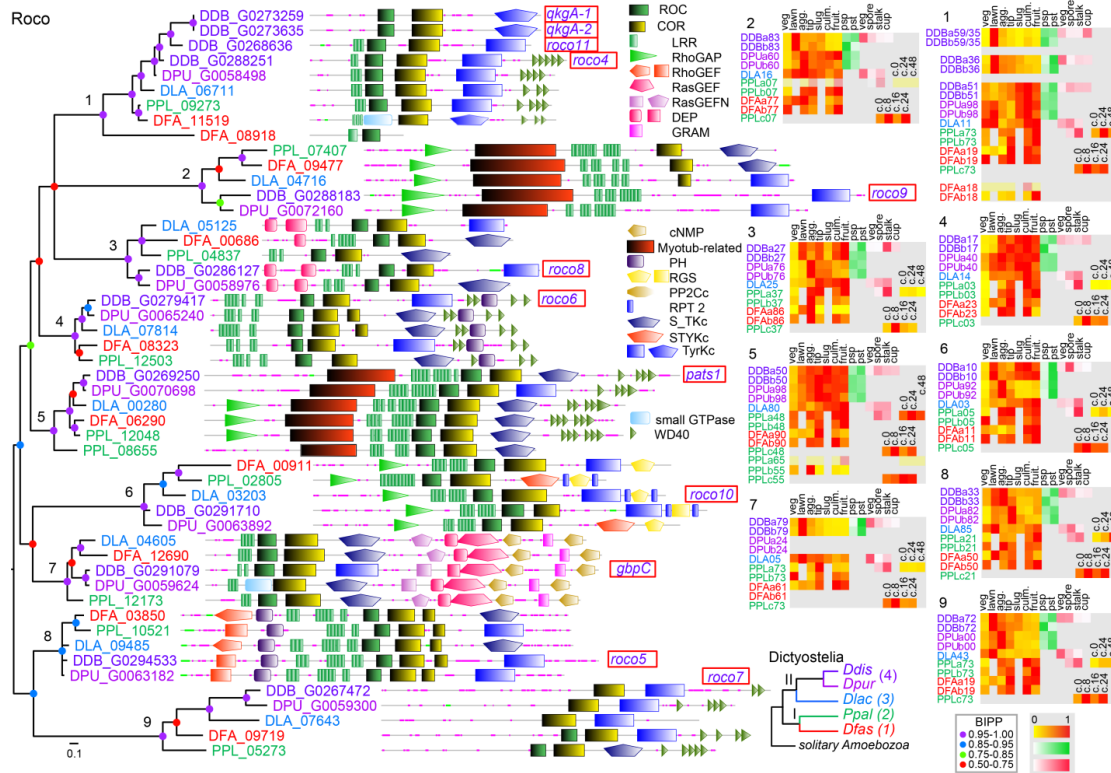


Figure 1. Conservation and change in the Roco GTPase family.

While most small GTPases only consist of the GTPase domain, the Roco proteins are unusual in combining a Roc (Ras Of Complex) GTPase domain with a COR (C-terminal of Roc) domain and a plethora of other domains. All GTPases in the *Ddis*, *Dpur*, *Dlac*, *Ppal* and *Dfas* genomes were identified from Interpro scans as outlined in Methods, and a pilot phylogenetic tree was prepared from the aligned isolated GTPase sequences. In this tree the Roco GTPases formed a single clade. The sequences in this clade were further supplemented with hits of BLASTp and tBLASTn queries of genomes using Roco sequences as bait. A final tree was then constructed from the Roc sequences by Bayesian inference (Ronquist and Huelsenbeck 2003), in which the colour of the gene ID matches that of the species shown in the Dictyostelid phylogeny (lower right). Bayesian posterior probabilities (BIPP) of the nodes are indicated by coloured dots.

The tree was annotated with gene names, which were framed in red for genes with known function and with the functional domain architecture of the proteins as analysed in SMART (Schultz et al. 1998). For overlapping domains, we selected the domain with the lowest E-value. The SMART (full colour) or PFAM (black-shaded rectangle) domain graphics and identifiers are listed in the figure and further domain information can be retrieved from http://smart.embl-heidelberg.de/smart/domain_table.cgi or <http://pfam.xfam.org/browse> using the identifier as bait. Clades of orthologous proteins and other groupings were further annotated with heatmaps of relative transcript levels at specific developmental stages or in specific cell types, which were retrieved from published RNA sequencing experiments (Gloeckner et al. 2016; Kin et al. 2018; Forbes et al. 2019; Parikh et al. 2010) (yellow-red: 0-1 fraction of maximum value), prespore or prestalk cells (white-green: 0-1 fraction of summed reads), or vegetative, spore, stalk and cup cells (white-red: 0-1 fraction of summed reads). Numbers preceded by c. represent hours of starvation in cells set up for encystation. Sets with maximally

10 or less reads are shown in wash-out color. Note that the phylogeny subdivides the GTPases in clades of conserved orthologs, with orthology further substantiated by similarity of domain architecture. Some *Ddis* genes such as *qkgA-1* and *qkgA-2* arose from a very recent duplication of part of chromosome 2. For such genes, transcripts were mapped to only one of the replicates, which is indicated by the last two digits of the locus tags of each gene, separated by /.

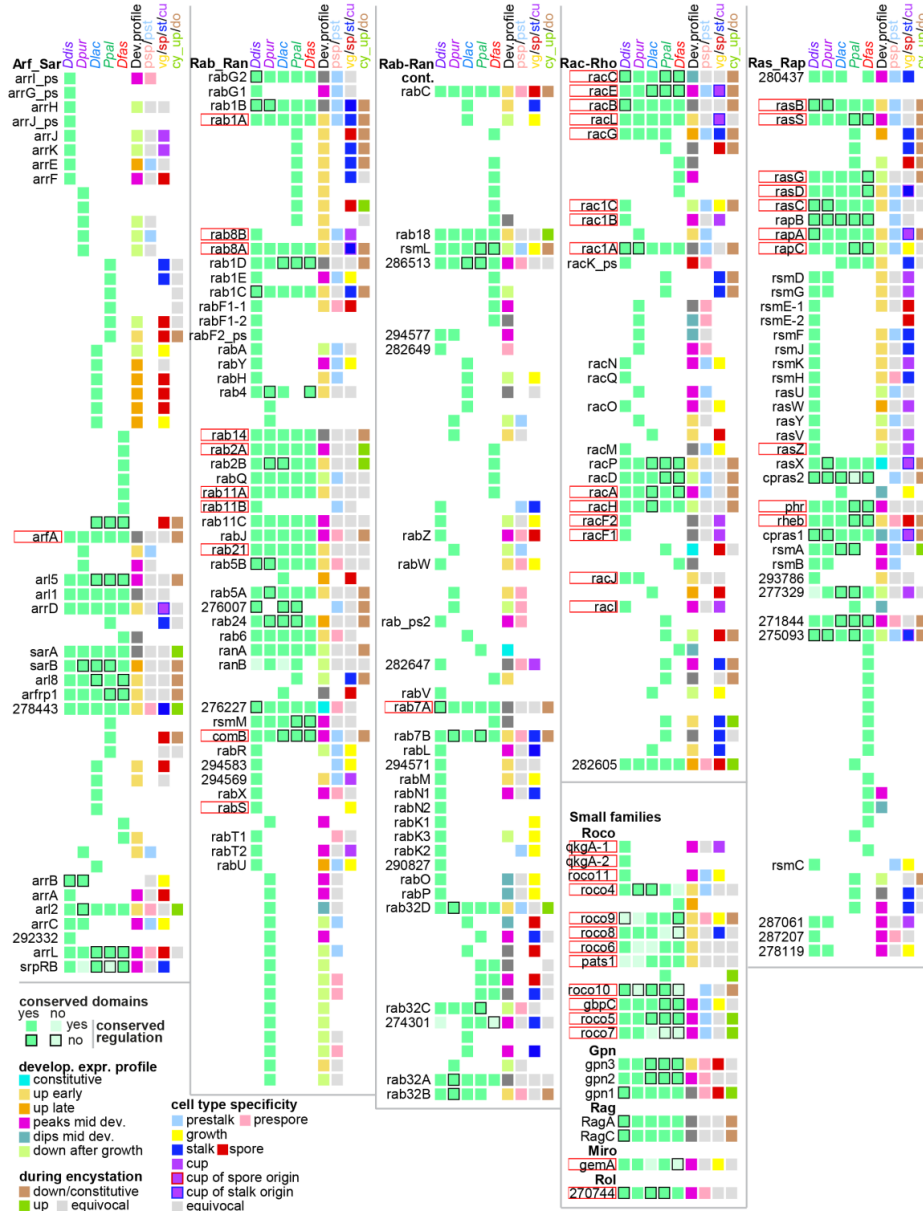


Figure 2. Summary of conservation and change in Dictyostelid GTPases

The presence of orthologous GTPases across the *Ddis*, *Dpur*, *Dlac*, *Ppal* and *Dfas* genomes is indicated by green squares in the first 5 columns, which are shown in pale green or with a black border, respectively, when compared to the majority, the functional domains or the developmental regulation are not conserved. Where the number of non-conserved features is larger than 3, pale green or a border is applied to all squares. The colour coding of the 6th, 7th and 8th square in each row respectively represent the developmental expression profile in the

majority of species, the prestalk/prespore specificity, when conserved between *Ddis* and *Dpur* slugs, the growth, spore or stalk specificity, when conserved between species, and the cup cell specificity in *Ddis*. The 9th square represents up- or down regulation in encystation of *Ppal*. Cup cells are only present in group 4 and are bordered red or blue when the orthologs in group 2 or 3 show spore- or stalk-specific expression, respectively. Grey reflects lack of specificity or conflicting data between species or replicate experiments, and white reflects absence of gene or data. The genes are listed by the *Ddis* gene names or 12 digit Dictybase gene identifiers without the DDB_G0 prefix. Genes with known function in *Ddis* are bordered in red. The gene identifiers and locus tags for the *Dpur*, *Dlac*, *Ppal* and *Dfas* genes are listed in Table S1.xlsx, together with all data on which this figure and figure 3 are based.

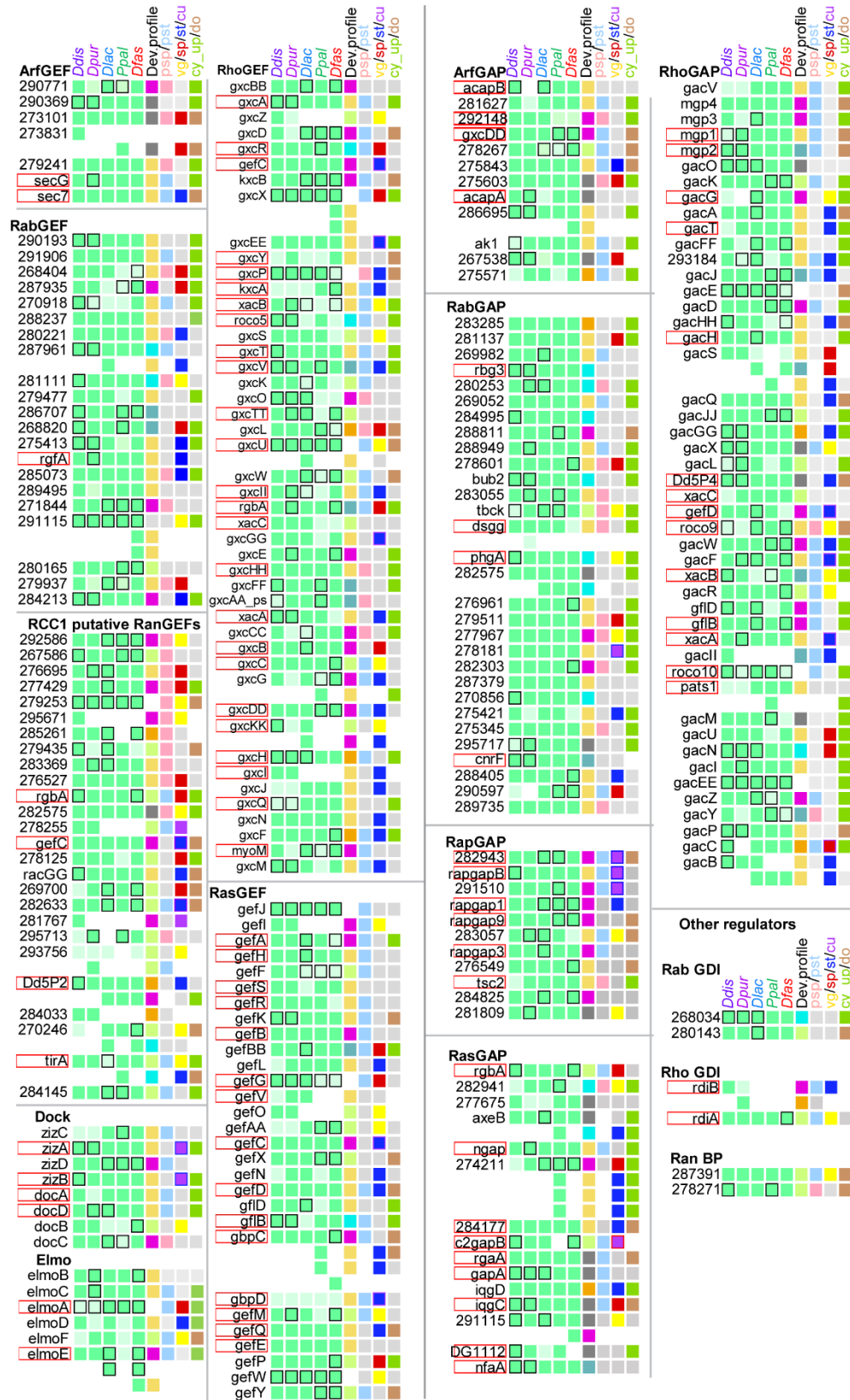


Figure 3. Summary of conservation and change in Dictyostelid GEFs, GAPs and other GTPase regulators.

Conservation and change in the presence, functional domain architecture and developmental regulation for sets of orthologous GEFs, GAPs and other regulators are summarized in the first five columns of squares, while the majority developmental profile, cell type specificity and regulation in *Ppal* encystation are shown in the next four columns. See figure 2 and its legend for further explanation. The gene identifiers and data upon which the figure is based are listed in Table S2.xlsx

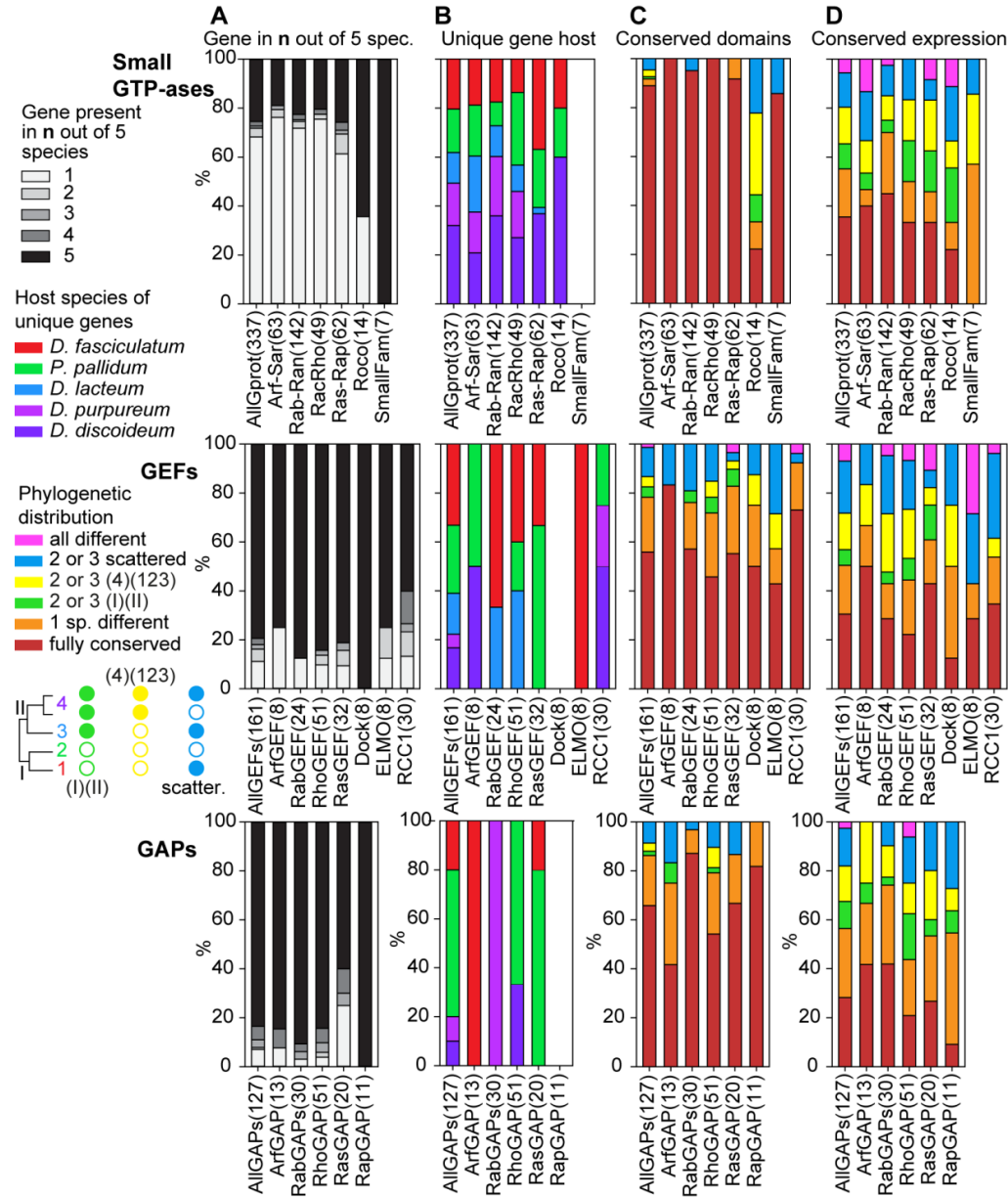


Figure 4. Conservation profiles of GTPases, GEFs and GAPs

For each GTPase, GEF or GAP family and all families of each category combined together, we calculated the percentage of the different states of the following features: A. the total number of orthologs out of five species that were conserved for each gene. B. The host species of genes that were unique. C/D. the phylogenetic distribution of genes with conserved domain and conserved regulation, respectively. The name of each family or grouping and its number of

members are shown at the X-axis. The figure is based on the data shown in supplemental figures S1-S16 that are compiled in supplemental tables S1 and S2.

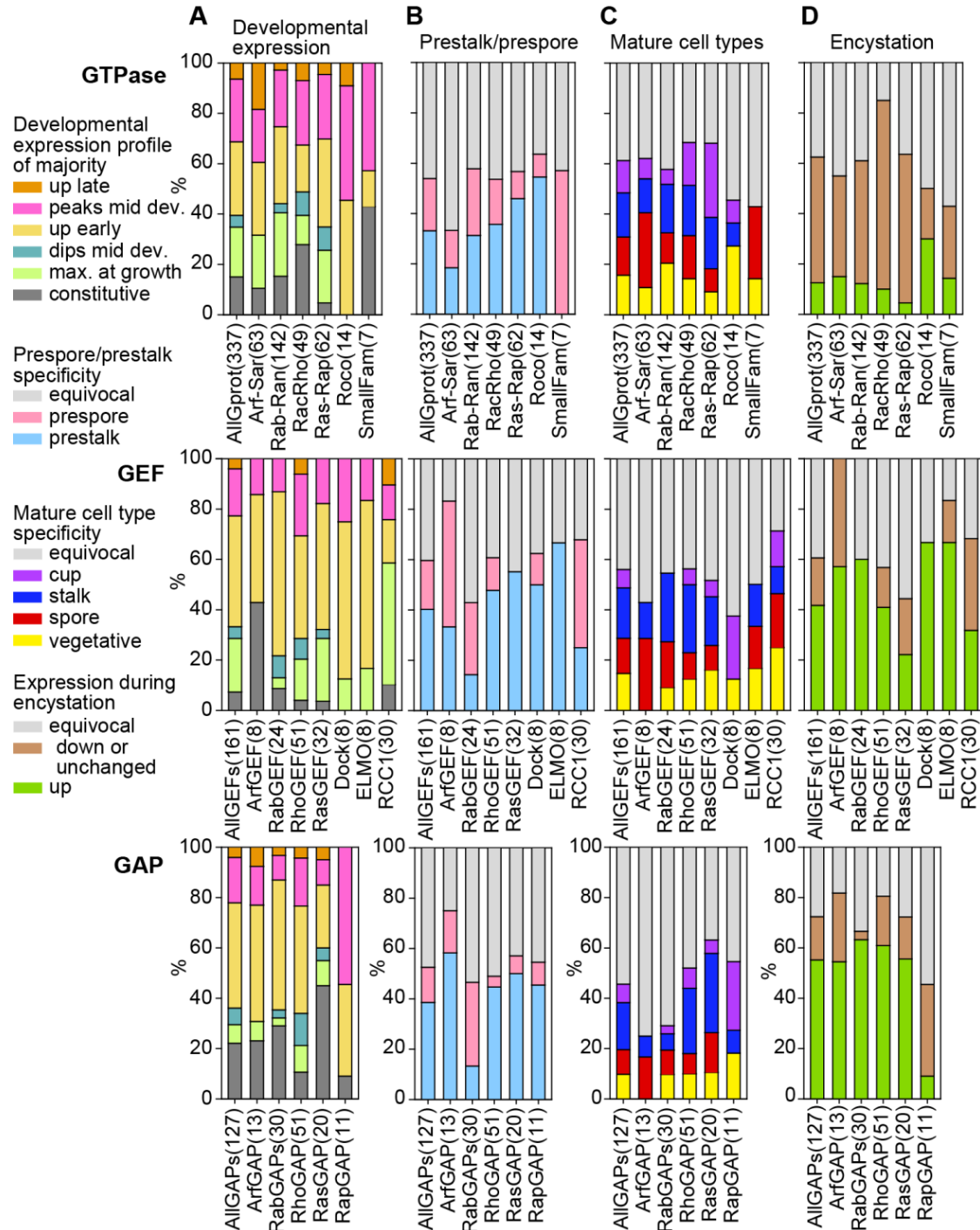


Figure 5. Developmental expression and cell-type specificity of GTPases, GEFs and GAPs.

For each GTPase, GEF or GAP family and all families of each category combined together, we calculated the percentage of the different states of the following features: A. The developmental expression profile of the majority of genes within orthologous groups. B. Prestalk

or prespore specificity in *Ddis* and/or *Dpur* slugs. C. Cell-type specificity in fruiting bodies of the majority of tested species (*Ddis*, *Dlac* and *Ppal*), compared to vegetative cells. C. Expression during encystation in *Ppal*.

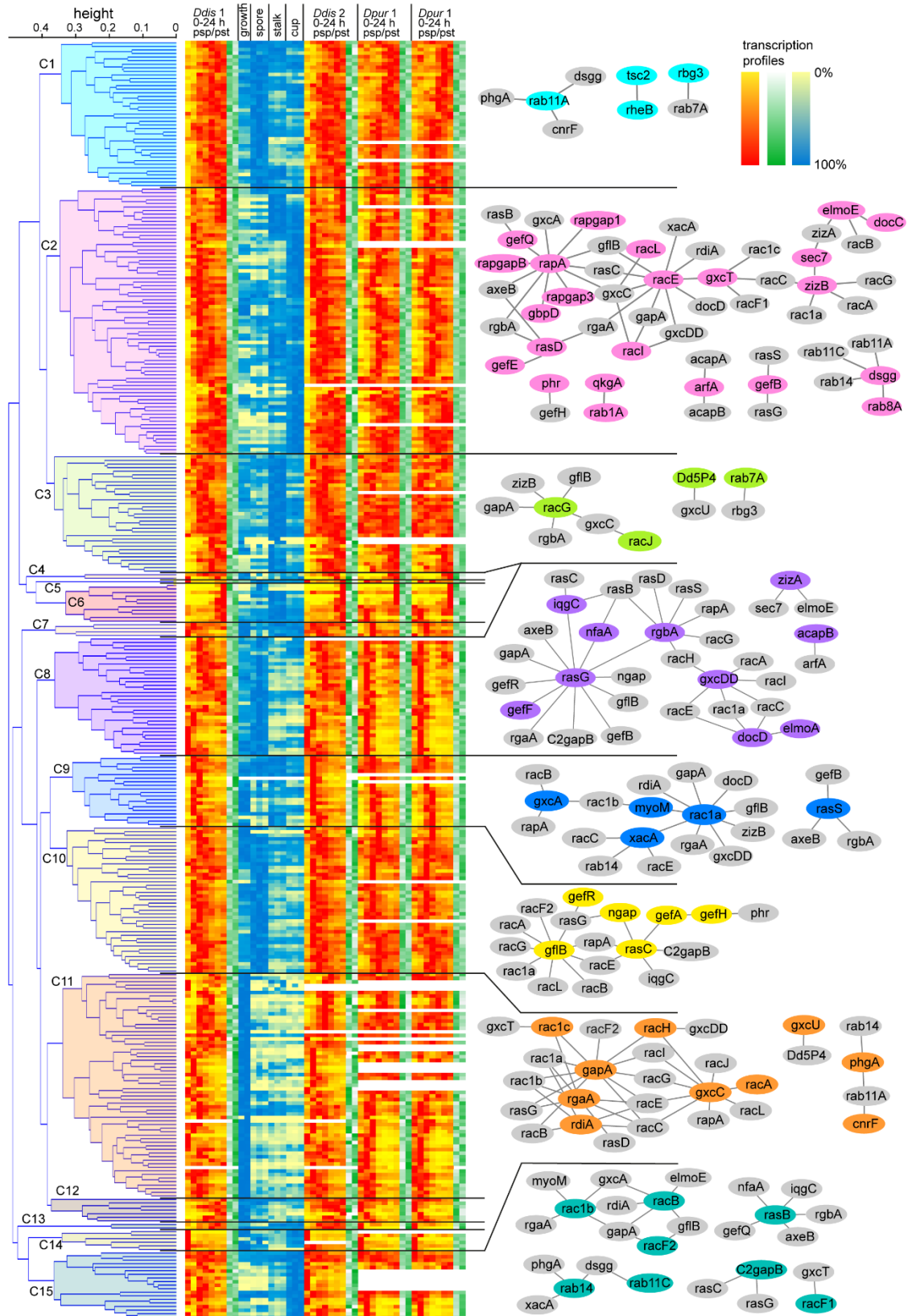


Figure 6. Hierarchical clustering of GTPases, GEFs and GAPs by expression profile.

Tree obtained by hierarchical clustering of developmental and cell-type specific gene expression data of GTPases and their direct regulators of the group 4 species *Ddis* and *Dpur*, using Pearson correlation to estimate distances between profiles and average linkage to infer the tree. Clusters (C) at a relative branch height of 75% are indicated and heatmaps of all clustered genes are shown. For clusters with >2 known interacting proteins as listed in Table S3.xlsx, sheet 4, the primary interacting partners inside (coloured) and outside (grey) the cluster are visualized as a network using Cytoscape (Shannon et al. 2003). The complete figure with individual genes annotated with locus tags, interactions, protein function and cellular localization is shown in Figure S17 and is derived from Table S4.xlsx, sheet "Group4". A network of all established GTPase and direct regulator interactions is shown in Figure S21.